CHAPTER - 3

A random variable Y is said to be discrete if it can assume only a finite or countably infinite number of distinct values

3.2 - The Probability Distribution for a Discrete Random Variable

Notation:

$Y \rightarrow$ random variable

$y \rightarrow$ particular value that a random variable may assume.

Example:

Let Y denote any one of 6 possibles values that could be observed on the upper face when a die is tossed. After the die is tossed, the number actually observed will be denoted by the symbol y

* Y is a random variable, but the specific observed value, y, is not random

* The expression $(Y=y)$ can be read as: the set of all points in S assigned the value y by the random variable Y.

$\therefore$ It is now meaningful to talk about the probability that Y takes on the value y, denoted by $P(Y=y)$

Definition: 3.2

The probability that Y takes on the value y, $P(Y=y)$ is defined as the sum of the probabilities of all sample points in S that are assigned the value y.

Lo denote by $P(Y=y) = \underline{p(y)}$

$\Rightarrow$ Where $\underline{p(y)}$ is a function that assigns probabilits to each value y of the random variable Y. [probability function for Y]

Definition 3.3

The probability distribution for a discrete variable Y can be represented by a formula, a table, or a graph. that provides $p(y) = P(Y = y)$ for all y.

Note:

$p(y) \geq 0$ for all y.

- but, the probability distribution for a discrete random variable assigns non-zero probabilities to only a countable number of distinct y-values
- Any value of y not explicitly assigned a positive probability is understood to be such that

p(y) = 0

EXAMPLE **3.1** A supervisor in a manufacturing plant has three men and three women working for him. He wants to choose two workers for a special job. Not wishing to show any biases in his selection, he decides to select the two workers at random. Let Y denote the number of women in his selection. Find the probability distribution for Y.

Solution The supervisor can select two workers from six in $\binom{6}{2} = 15$ ways. Hence, S contains 15 sample points, which we assume to be equally likely because random sampling was employed. Thus, $P(E_i) = 1/15$, for $i = 1, 2, \ldots, 15$. The values for Y that have nonzero probability are 0, 1, and 2. The number of ways of selecting $Y = 0$ women is $\binom{3}{0}\binom{3}{2}$ because the supervisor must select zero workers from the three women and two from the three men. Thus, there are $\binom{3}{0}\binom{3}{2} = 1 \cdot 3 = 3$ sample points in the event $Y = 0$, and

$$p(0) = P(Y = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = \frac{3}{15} = \frac{1}{5}.$$

Similarly,

$$p(1) = P(Y = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = \frac{9}{15} = \frac{3}{5},$$

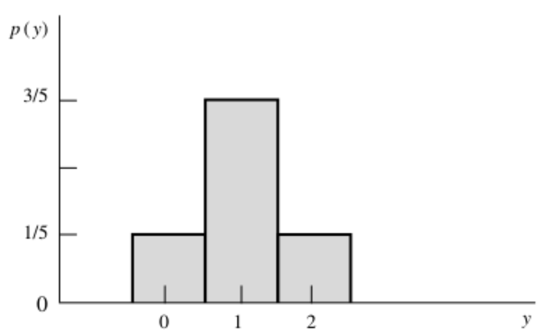
$$p(2) = P(Y = 2) = \frac{\binom{3}{2}\binom{3}{0}}{15} = \frac{3}{15} = \frac{1}{5}.$$

Notice that $(Y = 1)$ is by far the most likely outcome. This should seem reasonable since the number of women equals the number of men in the original group. ■

Table 3.1 Probability distribution for Example 3.1

y	$p(y)$
0	1/5
1	3/5
2	1/5

FIGURE **3.1**
Probability histogram for Table 3.1



Theorem 3.1

For any discrete probability distribution the following must be true:

1) $0 \leq p(y) \leq 1$ for all y.
2) $\sum_y p(y) = 1$, where the summation is al over all values of y with non-zero probability.

* As mentioned in Section 1-5, the <u>probability distributions</u> are models, <u>not</u> exact representations for the frequency distribution of populations of real data that occur (or would be generated) in nature

3.3 - The Expected Value of a Random Variable or Function of a Random Variable

- We have observed that the probability distribution for a random variable is a theoretical model for the empirical distribution of data associated with real population.
  - If the model is an accurate representation of nature, the theoretical & empirical distributions are equivalent

Definition 3.4

Let Y be a discrete random variable with the probability function $p(y)$. Then the expected value of Y, E(Y) is defined to be:

$$\boxed{E(Y) = \sum_y y\,p(y)}$$

↳ If p(y) is an accurate characterization of the population frequency distribution then $E(y) = \mu$, the population mean.

THEOREM 3.2

Let Y be a discrete random variable with probability function $p(y)$ and $g(Y)$ be a real-valued function of Y. Then the expected value of $g(Y)$ is given by

$$E[g(Y)] = \sum_{\text{all } y} g(y)p(y).$$

Proof

We prove the result in the case where the random variable Y takes on the finite number of values $y_1, y_2, \ldots, y_n$. Because the function $g(y)$ may not be one-to-one, suppose that $g(Y)$ takes on values $g_1, g_2, \ldots, g_m$ (where $m \leq n$). It follows that $g(Y)$ is a random variable such that for $i = 1, 2, \ldots, m$,

$$P[g(Y) = g_i] = \sum_{\substack{\text{all } y_j \text{ such that} \\ g(y_j) = g_i}} p(y_j) = p^*(g_i).$$

Thus, by Definition 3.4,

$$\begin{aligned}
E[g(Y)] &= \sum_{i=1}^{m} g_i p^*(g_i) \\
&= \sum_{i=1}^{m} g_i \left\{ \sum_{\substack{\text{all } y_j \text{ such that} \\ g(y_j) = g_i}} p(y_j) \right\} \\
&= \sum_{i=1}^{m} \sum_{\substack{\text{all } y_j \text{ such that} \\ g(y_j) = g_i}} g_i p(y_j) \\
&= \sum_{j=1}^{n} g(y_j) p(y_j).
\end{aligned}$$

→ Now, lets try to find numerical descriptive measures (parameters) to characterize p(y)

↳ E(Y) provides the mean of the population with distribution given by p(y)

Recall from Chapter 1

- Variance of a set of measurements is the average of the square of the differences between the values in a set of measurements and their means.

∴ $g(Y) = (Y - \mu)^2$

Definition. 3.5

If Y is a random variable with mean $E(Y) = \mu$, the variance of a random variable Y is defined to be the expected value of $(Y-\mu)^2$. That is:

$$\boxed{V(Y) = E[(Y-\mu)^2]}$$

* The standard deviation of Y is the positive square root of $V(Y)$

• Note •

If $p(y)$ is an accurate characterization of the population frequency distribution, then $E(Y) = \mu$, $V(Y) = \sigma^2$, the population variance, and σ is the population standard deviation.

* 3 useful expectation theorems that follow directly from the theory of summation. *

* Assume random variable Y with probability function p(y). *

1

THEOREM 3.3 Let Y be a discrete random variable with probability function $p(y)$ and c be a constant. Then $E(c) = c$.

Proof Consider the function $g(Y) \equiv c$. By Theorem 3.2,

$$E(c) = \sum_y cp(y) = c \sum_y p(y).$$

But $\sum_y p(y) = 1$ (Theorem 3.1) and, hence, $E(c) = c(1) = c$.

↳ States that the mean or expected value of a non-random quantity c is equal to c

2

THEOREM 3.4

Let Y be a discrete random variable with probability function $p(y)$, $g(Y)$ be a function of Y, and c be a constant. Then

$$E[cg(Y)] = cE[g(Y)].$$

Proof

By Theorem 3.2,

$$E[cg(Y)] = \sum_y cg(y)p(y) = c\sum_y g(y)p(y) = cE[g(Y)].$$

↳ states that the expected the expected value of the product of a constant & times a function of a random variable is equal to the constant times the expected value of the function of the variable

3

THEOREM 3.5

Let Y be a discrete random variable with probability function $p(y)$ and $g_1(Y)$, $g_2(Y), \ldots, g_k(Y)$ be k functions of Y. Then

$$E[g_1(Y) + g_2(Y) + \cdots + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + \cdots + E[g_k(Y)].$$

Proof

We will demonstrate the proof only for the case $k = 2$, but analogous steps will hold for any finite k. By Theorem 3.2,

$$\begin{aligned} E[g_1(Y) + g_2(Y)] &= \sum_y [g_1(y) + g_2(y)]p(y) \\ &= \sum_y g_1(y)p(y) + \sum_y g_2(y)p(y) \\ &= E[g_1(Y)] + E[g_2(Y)]. \end{aligned}$$

↳ states the mean or expected value of a sum of functions of a random variable Y is equal to the sum of their respective expected values

* 3.3, 3.4, 3.5 → can be used immediately to develop a theorem useful in finding the variance of a discrete random variable

4

THEOREM 3.6 Let Y be a discrete random variable with probability function $p(y)$ and mean $E(Y) = \mu$; then

$$V(Y) = \sigma^2 = E[(Y - \mu)^2] = E(Y^2) - \mu^2.$$

Proof

$$\begin{aligned}\sigma^2 &= E[(Y - \mu)^2] = E(Y^2 - 2\mu Y + \mu^2) \\ &= E(Y^2) - E(2\mu Y) + E(\mu^2) \qquad \text{(by Theorem 3.5).}\end{aligned}$$

Noting that μ is a constant and applying Theorems 3.4 and 3.3 to the second and third terms, respectively, we have

$$\sigma^2 = E(Y^2) - 2\mu E(Y) + \mu^2.$$

But $\mu = E(Y)$ and, therefore,

$$\sigma^2 = E(Y^2) - 2\mu^2 + \mu^2 = E(Y^2) - \mu^2.$$

Summary of section

Was to introduce the concept of an expected value and to develop useful theorems for finding means & variance of random variables or functions of random variables.

EXAMPLE 3.4 The manager of an industrial plant is planning to buy a new machine of either type A or type B. If t denotes the number of hours of daily operation, the number of daily repairs Y_1 required to maintain a machine of type A is a random variable with mean and variance both equal to $.10t$. The number of daily repairs Y_2 for a machine of type B is a random variable with mean and variance both equal to $.12t$. The daily cost of operating A is $C_A(t) = 10t + 30Y_1^2$; for B it is $C_B(t) = 8t + 30Y_2^2$. Assume that the repairs take negligible time and that each night the machines are tuned so that they operate essentially like new machines at the start of the next day. Which machine minimizes the expected daily cost if a workday consists of (a) 10 hours and (b) 20 hours?

Solution The expected daily cost for A is

$$E[C_A(t)] = E\left[10t + 30Y_1^2\right] = 10t + 30E\left(Y_1^2\right)$$
$$= 10t + 30\{V(Y_1) + [E(Y_1)]^2\} = 10t + 30[.10t + (.10t)^2]$$
$$= 13t + .3t^2.$$

In this calculation, we used the known values for $V(Y_1)$ and $E(Y_1)$ and the fact that $V(Y_1) = E(Y_1^2) - [E(Y_1)]^2$ to obtain that $E(Y_1^2) = V(Y_1) + [E(Y_1)]^2 = .10t + (.10t)^2$. Similarly,

$$E[C_B(t)] = E\left[8t + 30Y_2^2\right] = 8t + 30E\left(Y_2^2\right)$$
$$= 8t + 30\{V(Y_2) + [E(Y_2)]^2\} = 8t + 30[.12t + (.12t)^2]$$
$$= 11.6t + .432t^2.$$

Thus, for scenario (a) where $t = 10$,

$$E[C_A(10)] = 160 \quad \text{and} \quad E[C_B(10)] = 159.2,$$

which results in the choice of machine B.

For scenario (b), $t = 20$ and

$$E[C_A(20)] = 380 \quad \text{and} \quad E[C_B(20)] = 404.8,$$

resulting in the choice of machine A.

In conclusion, machines of type B are more economical for short time periods because of their smaller hourly operating cost. For long time periods, however, machines of type A are more economical because they tend to be repaired less frequently. ■

3.4 - The Binomial Probability Distr.

DEFINITION 3.6 A *binomial experiment* possesses the following properties:

1. The experiment consists of a fixed number, n, of identical trials.
2. Each trial results in one of two outcomes: success, S, or failure, F.
3. The probability of success on a single trial is equal to some value p and remains the same from trial to trial. The probability of a failure is equal to $q = (1 - p)$.
4. The trials are independent.
5. The random variable of interest is Y, the number of successes observed during the n trials.

DEFINITION 3.7 A random variable Y is said to have a *binomial distribution* based on n trials with success probability p if and only if

$$p(y) = \binom{n}{y} p^y q^{n-y}, \qquad y = 0, 1, 2, \ldots, n \text{ and } 0 \le p \le 1.$$

THEOREM 3.7 Let Y be a binomial random variable based on n trials and success probability p. Then

$$\mu = E(Y) = np \quad \text{and} \quad \sigma^2 = V(Y) = npq.$$

Proof By Definitions 3.4 and 3.7,

$$E(Y) = \sum_y y p(y) = \sum_{y=0}^{n} y \binom{n}{y} p^y q^{n-y}.$$

Notice that the first term in the sum is 0 and hence that

$$E(Y) = \sum_{y=1}^{n} y \frac{n!}{(n-y)!y!} p^y q^{n-y}$$

$$= \sum_{y=1}^{n} \frac{n!}{(n-y)!(y-1)!} p^y q^{n-y}.$$

The summands in this last expression bear a striking resemblance to binomial probabilities. In fact, if we factor np out of each term in the sum and let $z = y-1$,

$$E(Y) = np \sum_{y=1}^{n} \frac{(n-1)!}{(n-y)!(y-1)!} p^{y-1} q^{n-y}$$

$$= np \sum_{z=0}^{n-1} \frac{(n-1)!}{(n-1-z)!z!} p^z q^{n-1-z}$$

$$= np \sum_{z=0}^{n-1} \binom{n-1}{z} p^z q^{n-1-z}.$$

Notice that $p(z) = \binom{n-1}{z} p^z q^{n-1-z}$ is the binomial probability function based on $(n-1)$ trials. Thus, $\sum_z p(z) = 1$, and it follows that

$$\mu = E(Y) = np.$$

From Theorem 3.6, we know that $\sigma^2 = V(Y) = E(Y^2) - \mu^2$. Thus, σ^2 can be calculated if we find $E(Y^2)$. Finding $E(Y^2)$ directly is difficult because

$$E(Y^2) = \sum_{y=0}^{n} y^2 p(y) = \sum_{y=0}^{n} y^2 \binom{n}{y} p^y q^{n-y} = \sum_{y=0}^{n} y^2 \frac{n!}{y!(n-y)!} p^y q^{n-y}$$

and the quantity y^2 does not appear as a factor of $y!$. Where do we go from here? Notice that

$$E[Y(Y-1)] = E(Y^2 - Y) = E(Y^2) - E(Y)$$

and, therefore,

$$E(Y^2) = E[Y(Y-1)] + E(Y) = E[Y(Y-1)] + \mu.$$

In this case,

$$E[Y(Y-1)] = \sum_{y=0}^{n} y(y-1) \frac{n!}{y!(n-y)!} p^y q^{n-y}.$$

The first and second terms of this sum equal zero (when $y = 0$ and $y = 1$). Then

$$E[Y(Y-1)] = \sum_{y=2}^{n} \frac{n!}{(y-2)!(n-y)!} p^y q^{n-y}.$$

(Notice the cancellation that led to this last result. The anticipation of this cancellation is what actually motivated the consideration of $E[Y(Y-1)]$.) Again, the summands in the last expression look very much like binomial probabilities. Factor $n(n-1)p^2$ out of each term in the sum and let $z = y - 2$ to obtain

$$E[Y(Y-1)] = n(n-1)p^2 \sum_{y=2}^{n} \frac{(n-2)!}{(y-2)!(n-y)!} p^{y-2} q^{n-y}$$

$$= n(n-1)p^2 \sum_{z=0}^{n-2} \frac{(n-2)!}{z!(n-2-z)!} p^z q^{n-2-z}$$

$$= n(n-1)p^2 \sum_{z=0}^{n-2} \binom{n-2}{z} p^z q^{n-2-z}.$$

Again note that $p(z) = \binom{n-2}{z} p^z q^{n-2-z}$ is the binomial probability function based on $(n-2)$ trials. Then $\sum_{z=0}^{n-2} p(z) = 1$ (again using the device illustrated in the derivation of the mean) and

$$E[Y(Y-1)] = n(n-1)p^2.$$

Thus,

$$E(Y^2) = E[Y(Y-1)] + \mu = n(n-1)p^2 + np$$

and

$$\sigma^2 = E(Y^2) - \mu^2 = n(n-1)p^2 + np - n^2p^2$$

$$= np[(n-1)p + 1 - np] = np(1-p) = npq.$$

3.5 - The Geometric Probability Distr.

* Similar to binomial distribution, <u>but</u>:

→ Instead of the number of successes that occur in <u>n-trials</u> the geometric random variable Y. is the number of the trial on which the <u>first</u> success occurs.

* Thus, the experiment consist of a series of trials that concludes with the first success.

The sample space S for the experiment contains the countably infinite set of sample points:

$$\begin{array}{lll} E_1: & S & \text{(success on first trial)} \\ E_2: & FS & \text{(failure on first, success on second)} \\ E_3: & FFS & \text{(first success on the third trial)} \\ E_4: & FFFS & \text{(first success on the fourth trial)} \\ \vdots & & \\ E_k: & \underbrace{FFFF\ldots F}_{k-1}S & \text{(first success on the } k^{\text{th}} \text{ trial)} \\ \vdots & & \end{array}$$

Because the random variable Y is the number of trials up to and including the first success, the events $(Y = 1)$, $(Y = 2)$, and $(Y = 3)$ contain only the sample points E_1, E_2, and E_3, respectively. More generally, the numerical event $(Y = y)$ contains only E_y. Because the trials are independent, for any $y = 1, 2, 3, \ldots,$

$$p(y) = P(Y = y) = P(E_y) = P(\underbrace{FFFF\ldots F}_{y-1}S) = \underbrace{qqq\cdots q}_{y-1}\,p = q^{y-1}p.$$

DEFINITION 3.8

A random variable Y is said to have a *geometric probability distribution* if and only if

$$p(y) = q^{y-1}p, \qquad y = 1, 2, 3, \ldots, \quad 0 \le p \le 1.$$

Note: (Example of using geometric prob)

The geometric probability distribution is often used to model distributions of lengths of waiting times. For example, suppose that a commercial aircraft engine is serviced periodically so that its various parts are replaced at different points in time and hence are of varying ages. Then the probability of engine malfunction, p, during any randomly observed one-hour interval of operation might be the same as for any other one-hour interval. The length of time prior to engine malfunction is the number of one-hour intervals, Y, until the first malfunction. (For this application, engine malfunction in a given one-hour period is defined as a success. Notice that, as in the case of the binomial experiment, either of the two outcomes of a trial can be defined as a success. Again, a "success" is not necessarily what would be considered to be "good" in everyday conversation.)

Simple Example (3.11)

- Suppose that the probability of engine malfunction during any one-hour period is $p = 0.02$. Find the probability that a given engine will survive two hours.

→ Let Y denote the number of one-hour intervals until the first malfunction, we have:

$$P(\text{survive two hours}) = P(Y \geq 3)$$
$$= \sum_{y=3}^{\infty} p(y)$$

Because: $\sum_{y=1}^{\infty} p(y) = 1$

i.e.

$P(\text{survive two hours}) = 1 - \sum_{y=1}^{2} p(y)$

$= 1 - p - qp = 1 - 0.02 - (0.98)(0.02) = \underline{\underline{0.9604}}$

Note:

If we examine closer the geometric distribution given Definition 3.8 we can see that

→ Larger values of p (hence smaller values of q) lead to higher probabilities for the smaller values of Y.

& hence lower probabilities for the larger values of Y.

⇕

∴ The mean value of Y appears to be inversely proportional to p

* Proof of mean of random variable with geometric distribution

THEOREM 3.8 If Y is a random variable with a geometric distribution,

$$\mu = E(Y) = \frac{1}{p} \quad \text{and} \quad \sigma^2 = V(Y) = \frac{1-p}{p^2}.$$

Proof

$$E(Y) = \sum_{y=1}^{\infty} yq^{y-1}p = p\sum_{y=1}^{\infty} yq^{y-1}.$$

This series might seem to be difficult to sum directly. Actually, it can be summed easily if we take into account that, for $y \geq 1$,

$$\frac{d}{dq}(q^y) = yq^{y-1},$$

and, hence,

$$\frac{d}{dq}\left(\sum_{y=1}^{\infty} q^y\right) = \sum_{y=1}^{\infty} yq^{y-1}.$$

(The interchanging of derivative and sum here can be justified.) Substituting, we obtain

$$E(Y) = p\sum_{y=1}^{\infty} yq^{y-1} = p\frac{d}{dq}\left(\sum_{y=1}^{\infty} q^y\right).$$

The latter sum is the geometric series, $q + q^2 + q^3 + \cdots$, which is equal to $q/(1-q)$ (see Appendix $A1.11$). Therefore,

$$E(Y) = p\frac{d}{dq}\left(\frac{q}{1-q}\right) = p\left[\frac{1}{(1-q)^2}\right] = \frac{p}{p^2} = \frac{1}{p}.$$

To summarize, our approach is to express a series that cannot be summed directly as the derivative of a series for which the sum can be readily obtained. Once we evaluate the more easily handled series, we differentiate to complete the process.

The derivation of the variance is left as Exercise 3.85.

The next example, similar to Example 3.10, illustrates how knowledge of the geometric probability distribution can be used to estimate an unknown value of p, the probability of a success.

EXAMPLE 3.13 Suppose that we interview successive individuals working for the large company discussed in Example 3.10 and stop interviewing when we find the first person who likes the policy. If the fifth person interviewed is the first one who favors the new policy, find an estimate for p, the true but unknown proportion of employees who favor the new policy.

Solution If Y denotes the number of individuals interviewed until we find the first person who likes the new retirement plan, it is reasonable to conclude that Y has a geometric distribution for some value of p. Whatever the true value for p, we conclude that the probability of observing the first person in favor of the policy on the fifth trial is

$$P(Y = 5) = (1 - p)^4 p.$$

We will use as our estimate for p the value that maximizes the probability of observing the value that we *actually observed* (the first success on trial 5).

To find the value of p that maximizes $P(Y = 5)$, we again observe that the value of p that maximizes $P(Y = 5) = (1 - p)^4 p$ is the same as the value of p that maximizes $\ln[(1 - p)^4 p] = [4 \ln(1 - p) + \ln(p)]$.

If we take the derivative of $[4 \ln(1 - p) + \ln(p)]$ with respect to p, we obtain

$$\frac{d[4 \ln(1 - p) + \ln(p)]}{dp} = \frac{-4}{1 - p} + \frac{1}{p}.$$

Setting this derivative equal to 0 and solving, we obtain $p = 1/5$.

Because the second derivative of $[4 \ln(1 - p) + \ln(p)]$ is negative when $p = 1/5$, it follows that $[4 \ln(1 - p) + \ln(p)]$ [and $P(Y = 5)$] is *maximized* when $p = 1/5$. Our estimate for p, based on observing the first success on the fifth trial is $1/5$.

Perhaps this result is a little more surprising than the answer we obtained in Example 3.10 where we estimated p on the basis of observing 6 in favor of the new plan in a sample of size 20. Again, this is an example of the use of the *method of maximum likelihood* that will be studied in more detail in Chapter 9. ■

3.6 - The Negative Binomial Probability distribution (Optional)

⇝ Comes from the same origin as geometric distribution.

↳ What if we are interested in knowing the number of trials on which the second, third, or fourth success occurs?

↳ The distribution that applies to the random variable Y equal to the number of the trial on which the r^{th} success occurs ($r = 2, 3, 4$, etc..) → [negative binomial distribution]

- Let us select fixed values for r & y and consider events A & B where:

$A = \{$ the first $(y-1)$ trials contains $(r-1)$ successes$\}$

$B = \{$ trial y results in a success$\}$

- Because we assume that the trials are independent, it follows that A & $B \Rightarrow$ independent events.

(conditions)

$\therefore$

- $P(B) = p$

$\Rightarrow p(y) = p(Y=y) = P(A \cap B) = P(A) \times P(B)$

Notice that $P(A)$ is 0 if $(y-1) < (r-1)$ or, equivalently, if $y < r$. If $y \geq r$, our previous work with the binomial distribution implies that

$$P(A) = \binom{y-1}{r-1} p^{r-1} q^{y-r}.$$

Finally,

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r}, \qquad y = r, r+1, r+2, \ldots .$$

DEFINITION 3.9

A random variable Y is said to have a *negative binomial probability distribution* if and only if

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r}, \qquad y = r, r+1, r+2, \ldots, \quad 0 \leq p \leq 1.$$

THEOREM 3.9

If Y is a random variable with a negative binomial distribution,

$$\mu = E(Y) = \frac{r}{p} \quad \text{and} \quad \sigma^2 = V(Y) = \frac{r(1-p)}{p^2}.$$

EXAMPLE

1

EXAMPLE 3.14 A geological study indicates that an exploratory oil well drilled in a particular region should strike oil with probability .2. Find the probability that the third oil strike comes on the fifth well drilled.

Solution Assuming independent drillings and probability .2 of striking oil with any one well, let Y denote the number of the trial on which the third oil strike occurs. Then it is reasonable to assume that Y has a negative binomial distribution with $p = .2$. Because we are interested in $r = 3$ and $y = 5$,

$$P(Y = 5) = p(5) = \binom{4}{2}(.2)^3(.8)^2$$
$$= 6(.008)(.64) = .0307. \quad \blacksquare$$

②

EXAMPLE **3.15** A large stockpile of used pumps contains 20% that are in need of repair. A maintenance worker is sent to the stockpile with three repair kits. She selects pumps at random and tests them one at a time. If the pump works, she sets it aside for future use. However, if the pump does not work, she uses one of her repair kits on it. Suppose that it takes 10 minutes to test a pump that is in working condition and 30 minutes to test and repair a pump that does not work. Find the mean and variance of the total time it takes the maintenance worker to use her three repair kits.

Solution Let Y denote the number of the trial on which the third nonfunctioning pump is found. It follows that Y has a negative binomial distribution with $p = .2$. Thus, $E(Y) = 3/(.2) = 15$ and $V(Y) = 3(.8)/(.2)^2 = 60$. Because it takes an additional 20 minutes to repair each defective pump, the total time necessary to use the three kits is

$$T = 10Y + 3(20).$$

Using the result derived in Exercise 3.33, we see that

$$E(T) = 10E(Y) + 60 = 10(15) + 60 = 210$$

and

$$V(T) = 10^2 V(Y) = 100(60) = 6000.$$

Thus, the total time necessary to use all three kits has mean 210 and standard deviation $\sqrt{6000} = 77.46$. ■

3.7 - The Hypergeometric Probability Distribution

Reminder:

In example 3.6

- We concluded that if the sample size n was small relative to the population size

N, the distribution of Y can be approximated by a binomial distribution.

- If n is large relative to N, the conditional probability of selecting a supporter of Jones on a later draw would be significantly affected by the observed preferences of persons selected on earlier draws.

$\therefore$ Thus the trials were not independent and the probability distribution for Y could not be approximated adequately by a binomial probability distribution.

→ The hypergeometric probability can be derived by using the combinatorial theorem.

DEFINITION 3.10 A random variable Y is said to have a *hypergeometric probability distribution* if and only if

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}},$$

where y is an integer 0, 1, 2, ..., n, subject to the restrictions $y \leq r$ and $n - y \leq N - r$.

EXAMPLE 3.16 An important problem encountered by personnel directors and others faced with the selection of the best in a finite set of elements is exemplified by the following scenario. From a group of 20 Ph.D. engineers, 10 are randomly selected for employment. What is the probability that the 10 selected include all the 5 best engineers in the group of 20?

Solution For this example $N = 20$, $n = 10$, and $r = 5$. That is, there are only 5 in the set of 5 best engineers, and we seek the probability that $Y = 5$, where Y denotes the number

of best engineers among the ten selected. Then

$$p(5) = \frac{\binom{5}{5}\binom{15}{5}}{\binom{20}{10}} = \left(\frac{15!}{5!10!}\right)\left(\frac{10!10!}{20!}\right) = \frac{21}{1292} = .0162. \quad \blacksquare$$

THEOREM **3.10** If Y is a random variable with a hypergeometric distribution,

$$\mu = E(Y) = \frac{nr}{N} \quad \text{and} \quad \sigma^2 = V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

3.8 - Poisson Probability Distribution

DEFINITION **3.11** A random variable Y is said to have a *Poisson probability distribution* if and only if

$$p(y) = \frac{\lambda^y}{y!}e^{-\lambda}, \qquad y = 0, 1, 2, \ldots, \ \lambda > 0.$$

The Poisson probability distribution often provides a good model for the probability distribution of the number Y of rare events that occur in space, time, volume, or any other dimension, where λ is the average value of Y. As we have noted, it provides a good model for the probability distribution of the number Y of automobile accidents, industrial accidents, or other types of accidents in a given unit of time. Other examples of random variables with approximate Poisson distributions are the number of telephone calls handled by a switchboard in a time interval, the number of radioactive particles that decay in a particular time period, the number of errors a typist makes in typing a page, and the number of automobiles using a freeway access ramp in a ten-minute interval.

THEOREM **3.11** If Y is a random variable possessing a Poisson distribution with parameter λ, then

$$\mu = E(Y) = \lambda \quad \text{and} \quad \sigma^2 = V(Y) = \lambda.$$

Proof By definition,

$$E(Y) = \sum_y y p(y) = \sum_{y=0}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!}.$$

Notice that the first term in this sum is equal to 0 (when $y = 0$), and, hence,

$$E(Y) = \sum_{y=1}^{\infty} y \frac{\lambda^y e^{-\lambda}}{y!} = \sum_{y=1}^{\infty} \frac{\lambda^y e^{-\lambda}}{(y-1)!}.$$

As it stands, this quantity is not equal to the sum of the values of a probability function $p(y)$ over all values of y, but we can change it to the proper form by factoring λ out of the expression and letting $z = y - 1$. Then the limits of summation become $z = 0$ (when $y = 1$) and $z = \infty$ (when $y = \infty$), and

$$E(Y) = \lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1} e^{-\lambda}}{(y-1)!} = \lambda \sum_{z=0}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!}.$$

Notice that $p(z) = \lambda^z e^{-\lambda}/z!$ is the probability function for a Poisson random variable, and $\sum_{z=0}^{\infty} p(z) = 1$. Therefore, $E(Y) = \lambda$. Thus, the mean of a Poisson random variable is the single parameter λ that appears in the expression for the Poisson probability function.

We leave the derivation of the variance as Exercise 3.138.

3.9 - Moments and Moment - Generating Function

- The parameters μ & σ are meaningful numerical descriptive measures that locate the centre & describe the spread associated with the values of a random variable Y.

↳ They do not however provide a unique characterization of the distribution Y. ⇒ since many distributions possess the same means & standard deviations.

→ We now consider a set of numerical descriptive measures that (at least under certain conditions) uniquely determine $p(y)$

DEFINITION 3.12 The kth *moment of a random variable Y taken about the origin* is defined to be $E(Y^k)$ and is denoted by μ'_k.

DEFINITION 3.14 The *moment-generating function $m(t)$ for a random variable Y* is defined to be $m(t) = E(e^{tY})$. We say that a moment-generating function for Y exists if there exists a positive constant b such that $m(t)$ is finite for $|t| \le b$.

(More found in pg 163 – 166

3.10 – Probability-Generating Function (optional)

A mathematical device useful in finding the probability distributions and other properties of integer-valued random variables is the probability-generating function.

DEFINITION **3.15**

Let Y be an integer-valued random variable for which $P(Y = i) = p_i$, where $i = 0, 1, 2, \ldots$. The *probability-generating function* $P(t)$ for Y is defined to be

$$P(t) = E(t^Y) = p_0 + p_1 t + p_2 t^2 + \cdots = \sum_{i=0}^{\infty} p_i t^i$$

for all values of t such that $P(t)$ is finite.

Note:

An important class of discrete random variables is one in which Y represents a count and consequently takes integer values: $Y = 0, 1, 2, 3 \ldots$

Binomial, geometric, hypergeometric & Poisson random variables all fall in this class.

Because we already have the moment-generating function to assist in finding the moments of a random variable, of what value is $P(t)$? The answer is that it may be difficult to find $m(t)$ but much easier to find $P(t)$. Thus, $P(t)$ provides an additional tool for finding the moments of a random variable. It may or may not be useful in a given situation.

Finding the moments of a random variable is not the major use of the probability-generating function. Its primary application is in deriving the probability function (and hence the probability distribution) for other related integer-valued random variables. For these applications, see Feller (1968) and Parzen (1992).

3.11 - Tchebysheff's Theorem

→ can be used to determine a lower bound for the probability that the random variable Y of interest falls in an interval $\mu \pm k\sigma$

THEOREM 3.14 **Tchebysheff's Theorem** Let Y be a random variable with mean μ and finite variance σ^2. Then, for any constant $k > 0$,

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{or} \quad P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Important aspects

1. The result applies for any probability distribution (whether the probability histogram is bell-shaped or not.

2. The results of the theorem are very conservative in the sense that the actual probability that Y is in the interval $\mu \pm k\sigma$ usually exceeds the lower bound for the probability, $(1 - 1/k^2)$ by a considerable amount.

EXAMPLE **3.28** The number of customers per day at a sales counter, Y, has been observed for a long period of time and found to have mean 20 and standard deviation 2. The probability distribution of Y is not known. What can be said about the probability that, tomorrow, Y will be greater than 16 but less than 24?

Solution We want to find $P(16 < Y < 24)$. From Theorem 3.14 we know that, for any $k \geq 0$, $P(|Y - \mu| < k\sigma) \geq 1 - 1/k^2$, or

$$P[(\mu - k\sigma) < Y < (\mu + k\sigma)] \geq 1 - \frac{1}{k^2}.$$

Because $\mu = 20$ and $\sigma = 2$, it follows that $\mu - k\sigma = 16$ and $\mu + k\sigma = 24$ if $k = 2$. Thus,

$$P(16 < Y < 24) = P(\mu - 2\sigma < Y < \mu + 2\sigma) \geq 1 - \frac{1}{(2)^2} = \frac{3}{4}.$$

In other words, tomorrow's customer total will be between 16 and 24 with a fairly high probability (at least 3/4).

Notice that if σ were 1, k would be 4, and

$$P(16 < Y < 24) = P(\mu - 4\sigma < Y < \mu + 4\sigma) \geq 1 - \frac{1}{(4)^2} = \frac{15}{16}.$$

Thus, the value of σ has considerable effect on probabilities associated with intervals. ■

3.12 Summary

This chapter has explored discrete random variables, their probability distributions, and their expected values. Calculating the probability distribution for a discrete random variable requires the use of the probabilistic methods of Chapter 2 to evaluate the probabilities of numerical events. Probability functions, $p(y) = P(Y = y)$, were derived for binomial, geometric, negative binomial, hypergeometric, and Poisson random variables. These probability functions are sometimes called *probability mass functions* because they give the probability (mass) assigned to each of the finite or countably infinite possible values for these discrete random variables.

The expected values of random variables and functions of random variables provided a method for finding the mean and variance of Y and consequently measures of centrality and variation for $p(y)$. Much of the remaining material in the chapter was devoted to the techniques for acquiring expectations, which sometimes involved summing apparently intractable series. The techniques for obtaining closed-form expressions for some of the resulting expected values included (1) use of the fact that $\sum_y p(y) = 1$ for any discrete random variable and (2) $E(Y^2) = E[Y(Y-1)] + E(Y)$. The means and variances of several of the more common discrete distributions are summarized in Table 3.4. These results and more are also found in Table A2.1 in Appendix 2 and inside the back cover of this book.

Table 3.5 gives the R (and S-Plus) procedures that yield $p(y_0) = P(Y = y_0)$ and $P(Y \leq y_0)$ for random variables with binomial, geometric, negative binomial, hypergeometric, and Poisson distributions.

We then discussed the moment-generating function associated with a random variable. Although sometimes useful in finding μ and σ, the moment-generating function is of primary value to the theoretical statistician for deriving the probability distribution of a random variable. The moment-generating functions for most of the common random variables are found in Appendix 2 and inside the back cover of this book.

Table 3.4 Means and variances for some common discrete random variables

Distribution	$E(Y)$	$V(Y)$
Binomial	np	$np(1-p) = npq$
Geometric	$\frac{1}{p}$	$\frac{1-p}{p^2} = \frac{q}{p^2}$
Hypergeometric	$n\left(\frac{r}{N}\right)$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$
Poisson	λ	λ
Negative binomial	$\frac{r}{p}$	$\frac{r(1-p)}{p^2} = \frac{rq}{p^2}$

Table 3.5 *R* (and *S*-Plus) procedures giving probabilities for some common discrete distributions

Distribution	$P(Y = y_0) = p(y_0)$	$P(Y \leq y_0)$
Binomial	`dbinom(y0,n,p)`	`pbinom(y0,n,p)`
Geometric	`dgeom(y0-1,p)`	`pgeom(y0-1,p)`
Hypergeometric	`dhyper(y0,r,N-r,n)`	`phyper(y0,r,N-r,n)`
Poisson	`dpois(y0,λ)`	`ppois(y0,λ)`
Negative binomial	`dnbinom(y0-r,r,p)`	`pnbinom(y0-r,r,p)`

The probability-generating function is a useful device for deriving moments and probability distributions of integer-valued random variables.

Finally, we gave Tchebysheff's theorem a very useful result that permits approximating certain probabilities when only the mean and variance are known.

To conclude this summary, we recall the primary objective of statistics: to make an inference about a population based on information contained in a sample. Drawing the sample from the population is the experiment. The sample is often a set of measurements of one or more random variables, and it is the observed event resulting from a single repetition of the experiment. Finally, making the inference about the population requires knowledge of the probability of occurrence of the observed sample, which in turn requires knowledge of the probability distributions of the random variables that generated the sample.